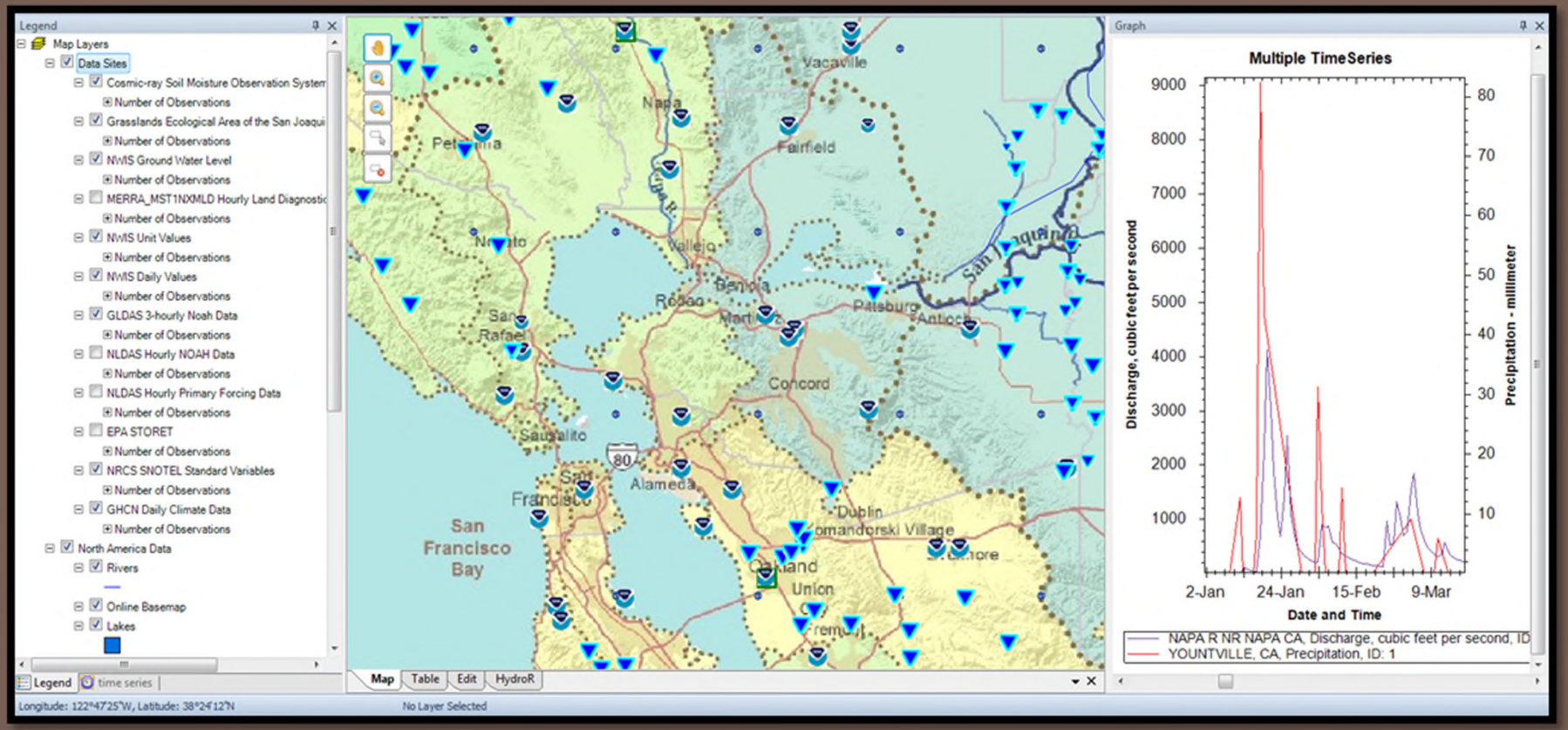


THE CUAHSI WATER DATA CENTER

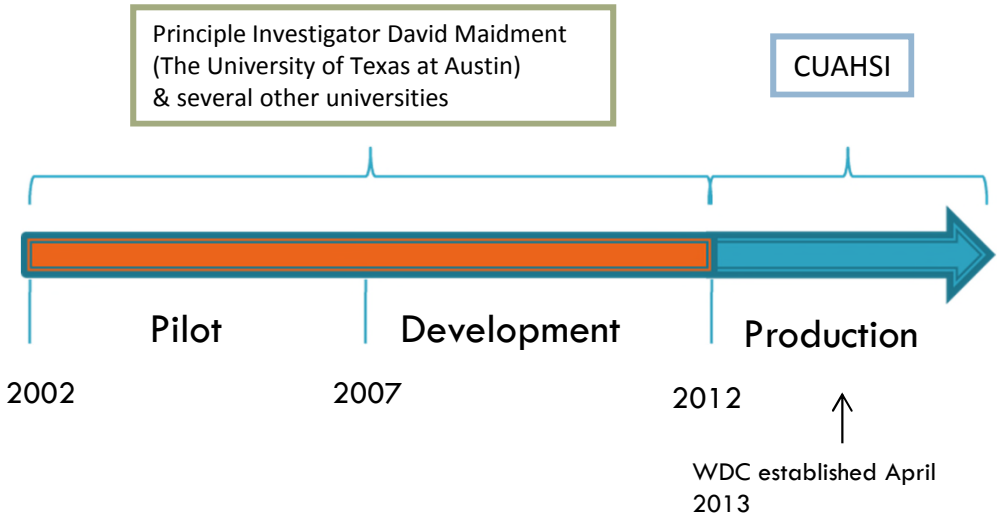
May 1, 2014

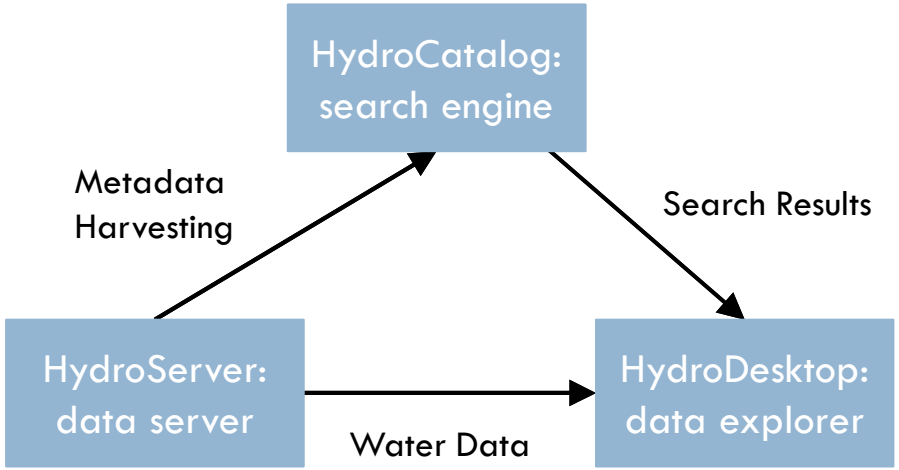
Jon Pollak, CUAHSI

Background

Consortium of Universities for the Advancement of Hydrologic Science, Inc.

- Established in 2001
- Over 100 members

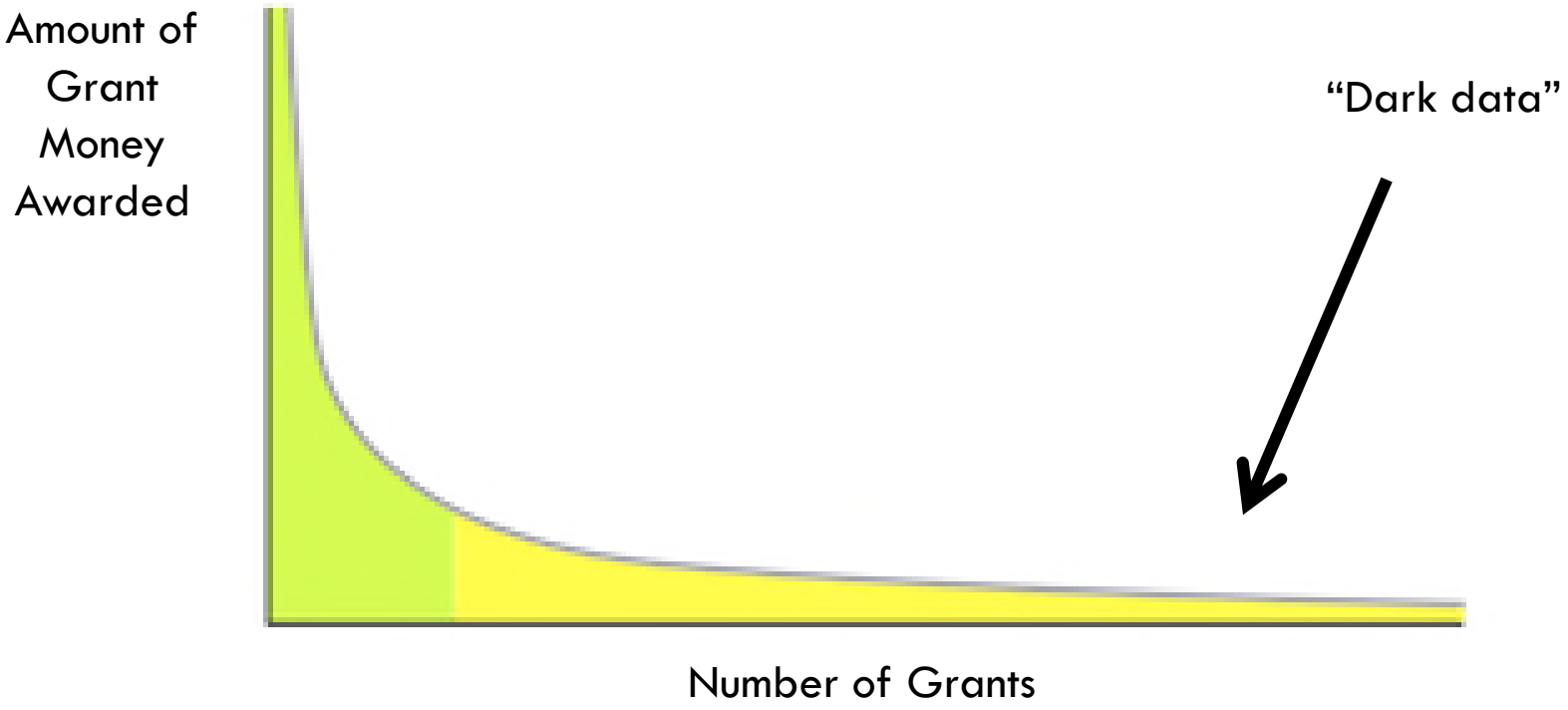




Mission

- Provide data services for the academic community
  - Data Management/Archive
  - Data Access
- To define how to achieve this mission, CUAHSI uses a community governance structure that directs staff:
  - Board of Directors
  - Standing Committee on Informatics
  - WDC Users Committee

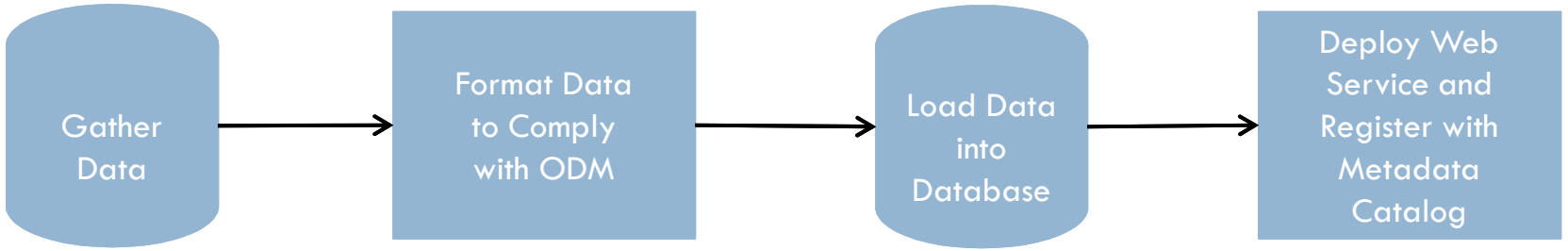
The "Long Tail" of Science



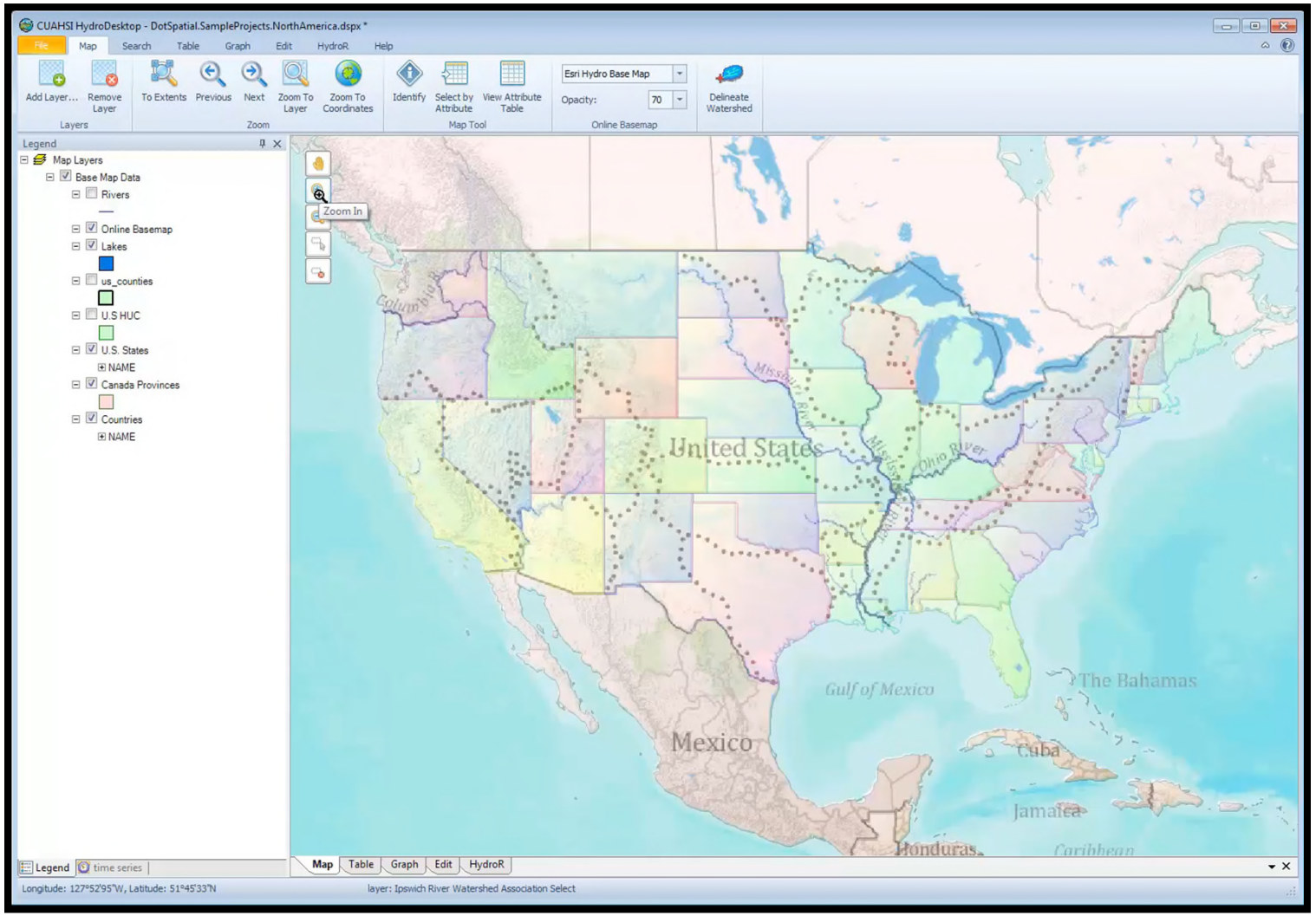
Heidorn (2008). *Shedding Light on the Dark Data in the Long Tail of Science*. Library Trends, *57*(2).

Benefits for Data Publication

- Adherence to Standards (ODM1 → ODM2; WaterML1 → WaterML2)
- Free and Open Source Software, Cloud Storage
- We are funded to deal with issues of long term data persistence



Benefits for Data Access



User-Centered Engineering Around Standards

Initially…	One year in…
CUAHSI HIS is **software**	CUAHSI HIS is a **service**
Most important parts of CUAHSI HIS are its **software components.**	Most important parts of CUAHSI HIS are its **standards and best practices.**
Addressing problems requires **repairing existing software**	Addressing problems requires **user-centered (agile) engineering**
A software ecosystem based upon **enabling individual research achievement**	A software ecosystem based upon **enabling global water science**

- **There is enormous utility in de-facto standards;** these create the leverage that evolves true global standards.
- **Standards are more important to maintain and nurture** than software.

Challenges

- Existing brittle infrastructure→ User-centered re-engineering
- Quality control of data→ Develop ways of expressing and measuring data quality
- Semantic mediation→ Develop approaches to cross-domain data discovery, including controlled vocabularies
- Data curation & metadata compliance→ Develop tools that assist users in specifying metadata
- Rapidly evolving data formats and standards→ Continuous re-engineering
- The cost of persistence→ Economical and near-line data storage alternatives
- New kinds of time series→ New metadata and catalog structures

Thank you! Questions?